

Center
for
Economic Research

No. 2000-45

**SEXUAL BIAS AND HOUSEHOLD
CONSUMPTION: A SEMIPARAMETRIC
ANALYSIS OF ENGEL CURVES IN RURAL
CHINA**

By Xiaodong Gong, Arthur van Soest and Ping Zhang

April 2000

ISSN 0924-7815

Sexual Bias and Household Consumption: A Semiparametric Analysis of Engel Curves in Rural China

Xiaodong Gong*, Arthur van Soest[†] and Ping Zhang[‡]
March 2000

Abstract

We analyze Engel curves for nuclear households in rural China. The sample includes more than 5000 nuclear families covering nineteen out of thirty Chinese provinces. We consider expenditures on food, also subdivided into several food subcategories such as cereals, or meat and fish, and other consumption categories such as alcohol and tobacco, medical, and educational goods. We use the semiparametric partially linear model. This allows for any functional form relationship between the budget shares and total expenditures, but assumes that the demographic variables enter the model in a linear way. We correct for potential endogeneity of total expenditures. Our results suggest that there are economies of scale in families' consumption expenditure patterns. We find some differences in consumption patterns which relate to differences in gender of children, which can be seen as evidence of sexual bias related to a commonly believed existing preference for boys.

*Tilburg University, P.O. Box 90153, 5000LE, Tilburg, the Netherlands, e-mail: x.gong@kub.nl

[†]Tilburg University

[‡]Institute of Economics, the Chinese Academy of Social Studies

The authors thank Bertrand Melenberg for helpful suggestions.

1 Introduction

Evaluating the welfare of the poor and analyzing the impact of income policy and demographic changes on the welfare of the poor are of main concern in most societies. For this purpose, recovering the exact Engel curve relationship is an important tool. For example, the budget share spent on food is an important indicator of household welfare. The rate at which a rise in income generates a decline in this budget share depends on the shape of the Engel curve of food. Engel curves for families of different composition can be used to reveal how family composition affects the food share, keeping family income constant. Differences in expenditures on typical adult goods can be used to analyze differentials in the allocation of family budget to the needs of boys versus girls (Deaton, 1989).

In this paper, we analyze Engel curves for several commodities in rural China using semiparametric methods. The aim is twofold. The first is to examine the shape of the Engel curves, which is relevant for the analysis of consumer demand and welfare. There are few studies on Engel curves in China. Fan *et al.* (1994, 1995) estimate Engel curves using data aggregated at the provincial level. Following most of the papers in this field, they used specifications in which the budget shares are linear in log income (linear Engel curves), in line with, for example, the Almost Ideal Demand System developed by Deaton and Muellbauer (1980). Recent evidence, however, shows that a linear Engel curve might not be appropriate for many commodities. See, for example, Atkinson *et al.* (1990), Hausman *et al.* (1991, 1995), Härdle and Jerison (1991), Lewbel (1991), Blundell *et al.* (1998), and Bhalotra and Attfield (1998). Moreover, analyses based upon aggregated data may be misleading if relationships are nonlinear (see Blundell *et al.*, 1993). Aggregate data cannot be used to estimate the impact of demographic changes or to calculate the equivalence scales.

Second, we focus on differences between consumption patterns of households with boys versus households with girls in various age groups. Thus, our second purpose is to determine whether the commonly found boy preference in China is embodied in household consumption behavior. In China, families have preferred boys over girls for centuries. But just as in other countries, evidence for this is largely based on census data on gender differentials in mortality and school enrollment rates.¹ See Arnold and Liu (1986), Zeng (1988), and Zeng *et al.* (1993). Do household consumption patterns reflect this preference? Empirical studies for various countries other than China do not find much evidence for this. See, for example, Deaton (1989) on Cote d'Ivoire and Thailand, Browning and Subramaniam (1995), Subramaniam (1996) and Subramaniam and Deaton (1991) on India, Ahmad and Morduch (1993) on Bangladesh, and Bhalotra and Attfield (1998) on Pakistan. To our knowledge, the only study which aims at identifying the boy-girl differentials in the consumption pattern of Chinese households

¹It has been found in the literature as evidence of sexual bias that in many low income countries the morbidity and mortality rates are higher for females, especially among children. See, for example, D'Souza and Chen (1980), and Das Gupta (1987).

using micro data is Burgess and Zhuang (1998). They use data on only two out of the thirty Chinese provinces. Their analysis is based upon the Working-Leser framework of linear Engel curves. If this functional form relationship is misspecified, this may also affect the coefficients on the demographics and the conclusions on sexual biases. They find some evidence of boy preference in household expenditures on health and on education.

Our paper differs from the Burgess and Zhuang (1998) study in the following ways. First, our data set is more representative of rural China as a whole. Second, we avoid the misspecification of the Engel curves by using a flexible semiparametric specification. Following Bhalotra and Attfield (1998), we apply the estimator of the partially linear regression model introduced by Robinson (1988). Third, since total expenditures might well be determined simultaneously with the expenditure shares, we allow for the endogeneity of the total expenditures. Here we follow Blundell *et al.* (1998), applying the method developed by Holly and Sargan (1982).

The finding suggests that linear Engel curves are suitable for some but not all of the nine expenditure categories we considered. For example, we find that the food expenditure share starts decreasing only after some threshold level of total expenditures is reached. It also suggests that failure in correcting for endogeneity of total expenditures leads to substantial biases in some of the elasticity estimates. Hardly any evidence for boy preference is found from food consumption. However, both Probits for school enrollment and analysis on educational expenditures conditional on enrollment provide clear evidence for biases toward boys.

Our paper is organized as follows. In Section 2, we present the semiparametric model and the estimation techniques. Section 3 describes the data used in the analysis and addresses preliminary evidence of gender bias which can be detected in our data. In Section 4, we discuss the results: the semiparametric estimates of the Engel curves are presented in Section 4.1, while the implications for gender bias are addressed in Section 4.2. Section 5 concludes.

2 The Model

The general form of the Engel curve is given by

$$y_i = g_i(\ln c, x) + \epsilon_i \quad (1)$$

where y_i represents the expenditure share on some commodity or group of commodities i , $\ln c$ is the log transformed total expenditures, x is a vector of variables that characterize family composition, and ϵ_i stands for the error term. The index of the individual household is suppressed. It is assumed that g is common to all households, so that variation across households with the same total expenditures c and the same composition x is only due to the error term ϵ_i . For the time being, we will assume that the error term

satisfies

$$E(\epsilon_i \mid \text{lnc}, x) = 0 \quad (2)$$

The more general case where ϵ is allowed to be correlated with lnc is discussed below.

In x , we include variables such as *the number of young children as a percentage of family size* and *the number of young girls as a percentage of family size*. The former is the sum of the latter and the *the number of young boys as a percentage of family size*. Thus, a test of significance of the *percentage of young girls* variable will be a test of whether the sex of a young child affects the consumption of good i or not, controlling for, among other variables, the number of young children.

Until recently, stemming from Working (1943) and Leser (1963), it has been popular to specify $g(\text{lnc}, x)$ parametrically, and in particular, as a linear function. Thus, it has been common to work with the specification

$$y_i = \alpha_i + \gamma_i \text{lnc} + \beta'_i x + \epsilon_i \text{ with } E(\epsilon_i \mid \text{lnc}) = 0. \quad (3)$$

This specification can be derived from, for example, the Almost Ideal Demand System by Deaton and Muellbauer (1980), or the Indirect Translog model by Jorgenson *et al.* (1980). It is also referred to as Piglog.

The linearity assumption might be too restrictive for some goods. For example, Blundell *et al.* (1998) found that, for their UK data, the Piglog specification fits the Engel curve for food quite well, but for a commodity like alcohol, adding a quadratic logarithmic income term is required.

An alternative to the linearity assumption is to find the correct specification using a fully nonparametric approach. A fully nonparametric model has the advantage that misspecification is by definition avoided. The approach, however, becomes infeasible in practice if the dimension of x is large and the number of observations is limited. Nonparametric estimators then suffer from the curse of dimensionality: due to the slow rate of convergence of the estimator, the estimates will not be accurate in finite samples of the size typically available (see Härdle, 1990, for example).

In our case, we want to work with rather detailed indicators of the presence of boys and girls in various age groups, so that the dimension of x will be rather large. Considering the size of our sample of about 5400 observations, this makes a completely nonparametric estimator infeasible.

Robinson's (1988) semiparametric partially linear model allows for a flexible functional form relationship between lnc and y_i , but at the same time avoids the curse of dimensionality. It gives the following specification of the Engel curve:

$$y_i = \beta'_i x + F_i(\text{lnc}) + \epsilon_i \quad (4)$$

The nonparametric part of this model is the unknown function F_i . This is a function of one variable only, and thus the curse of dimensionality is avoided. The family composition variables enter through the parametric part $\beta'_i x$. This has the advantage that a

test for sexual bias in consumption patterns can be written as a test on restrictions on the parameter vector β_i and does not involve testing restrictions on the nonparametric part F_i .

Estimation

Following Robinson (1988), the partially linear regression model can be estimated in various steps. The expected value of y_i conditional on lnc is given by

$$E(y_i | \text{lnc}) = \beta_i' E(x | \text{lnc}) + F_i(\text{lnc}) \quad (5)$$

The difference between (4) and (5) yields

$$y_i - E(y_i | \text{lnc}) = \beta_i'(x - E(x | \text{lnc})) + \epsilon_i \quad (6)$$

In the first step of Robinson's estimation procedure, the conditional means $E(y_i | \text{lnc})$ and $E(x | \text{lnc})$ are estimated nonparametrically, using one dimensional kernel regressions. In the second step, the conditional means being replaced by their nonparametric estimates, β_i in (6) is estimated by OLS. Robinson (1988) shows that the resulting estimator for β_i is \sqrt{n} consistent and asymptotically normal.

There are several ways to obtain consistent estimates $\hat{F}_i(\text{lnc})$ for the values of the unknown function F_i . One consistent estimator is based upon:

$$\hat{F}_i(\text{lnc}) = \hat{E}(y_i | \text{lnc}) - \hat{\beta}_i' \hat{E}(x | \text{lnc}) \quad (7)$$

where $\hat{E}(y_i | \text{lnc})$ and $\hat{E}(x | \text{lnc})$ are the nonparametric estimates of $E(y_i | \text{lnc})$ and $E(x | \text{lnc})$, which were already used to estimate β_i . Robinson (1988) shows that the fact that β_i is estimated does not matter for the asymptotic distribution of $\hat{F}_i(\text{lnc})$, which is therefore the same as that of $\hat{E}(y_i | \text{lnc}) - \beta_i' \hat{E}(x | \text{lnc})$.

A second consistent estimator for the values of \hat{F}_i is given by

$$\hat{F}_i(\text{lnc}) = \hat{E}(y_i - \hat{\beta}_i' x | \text{lnc}) \quad (8)$$

where the right-hand side is a nonparametric regression estimate of $y_i - \hat{\beta}_i' x$ on lnc . In this study, we used this second method, since it will make it easier to calculate the elasticities. It only requires one nonparametric regression instead of $k + 1$ in the case of (7).

Elasticities

The elasticity of expenditures on commodity i with respect to the total expenditures is one plus the elasticity of the expenditure share with respect to the total expenditures. For each household, the elasticity of expenditures on commodity i is calculated as the

percentage change of the expenditures on good i corresponding to a one-percent increase of the household's total expenditures. Given the consistent estimates $\hat{F}_i(\text{lnc})$ and $\hat{\beta}_i$, for each household, the expenditures on good i are computed at the actual total expenditures and at the one percent increased total expenditures, respectively. There are two ways to combine the household specific elasticities into an aggregate elasticity. One is to compute the (unweighted) average elasticity across all households. The other is to take the weighted average, using households' expenditures on commodity i as weights. We use the latter, which answers the question by how much expenditures on commodity i will change if total expenditures of all households increase by one percent. This seems the more relevant concept from a policy point of view. The 95% confidence interval is also computed, using a bootstrap method: the data set was resampled randomly with replacement, the parameters and the elasticities were estimated for each resampled data set. And the bounds of the confidence interval were computed from the standard deviations of the elasticities series.

Endogeneity of Total Expenditures

Until now we have assumed that the error term has conditional mean zero given total expenditures lnc (and other variables x). This assumption may be problematic due to potential correlation between lnc and ϵ_i , see, for example, Blundell *et al.* (1998). If $E(\epsilon_i \mid \text{lnc}, x) \neq 0$, the estimators discussed above will not be consistent for the parameters of interest.

To take endogeneity of lnc into account, we follow Blundell *et al.* (1998) and use the popular augmented regression technique introduced by Holly and Sargan (1982). Assume that there exists a vector z such that

$$\text{lnc} = \pi'z + \zeta \quad (9)$$

with

$$E(\zeta \mid z, x) = 0 \quad (10)$$

and

$$E(\epsilon_i \mid \text{lnc}, z, x) = \rho_i \zeta \quad (11)$$

Then we can write

$$y_i = \beta_i'x + F_i(\text{lnc}) + \rho_i \zeta + \tilde{\epsilon}_i \quad (12)$$

with

$$E(\tilde{\epsilon}_i \mid \text{lnc}, x, z) = 0 \quad (13)$$

If ζ were observed, this would lead to the same partially linear regression as before, with regressors $\tilde{x} = (x, \zeta)$ instead of x and parameters $\tilde{\beta}_i = (\beta_i, \rho_i)$ instead of β . Thus β_i and ρ_i could be estimated using the Robinson (1988) estimator described above.

Since ζ is not observed, however, an additional preliminary step is required: OLS of lnc on z yields OLS residuals $\hat{\zeta}$, which are consistent estimates of ζ . The semiparametric

estimator discussed above will then be applied with ζ replaced by $\hat{\zeta}$. Identification requires that the vector of instruments z contains at least one instrument which is not in x . We use the log of disposable family income for this. Moreover, z will also include some variables already included in x : regional dummies, average age of the couple, family size, and family composition indicators.

Smoothness Parameters

For the nonparametric regressions, we need to choose the kernels and the smoothness parameters. Throughout the analysis, we will use a quartic kernel function. We adopt the method of “Generalized Cross Validation” (GCV) (see Craven and Wahba, (1979)) to choose the bandwidths for all of the nonparametric regressions. GCV has the same asymptotic optimality properties as the complete cross validation procedure, but is computationally much more convenient. For arbitrary bandwidth h , the estimates \hat{y}_i (written as an $n \times 1$ vector, where n is the number of observations, are a linear function of (the $n \times 1$ vector) y_i , with coefficients depending on the bandwidth h and the observations of the explanatory variables, stored in a matrix X . They can thus be written as $\hat{y}_i = A(X, h)y_i$. GCV then consists of choosing the value h_i of h which minimizes

$$1/n\|[I - A(X, h)]y\|^2 / \{1/n \text{ trace}[I - A(X, h)]\}^2 \quad (14)$$

It can be shown that this choice of h is asymptotically optimal, in the sense that it minimizes the expected quadratic loss $E[(y_i - \hat{y}_i)^2]$ (see Newey *et al.* (1990), for example).

3 Data

The data we use were drawn from the “Rural Household Income and Expenditure Survey” conducted by the State Statistics Bureau of China (SSB) and the Chinese Academy of Social Science (CASS). This survey is comparable to the national household survey conducted by SSB; the respondents are a subsample of the national household survey, and the questionnaire is similar.² Our survey was drawn in 1995 and provides information on 7998 households in rural areas of 19 Chinese provinces.³ It contains detailed questions on income, expenditures, and self-consumption, on land and capital ownership and financial assets, on labor market status of the household members, etc. It contains age, gender, and relation with the head of household for all members. For household members of at least 15 years of age, it includes variables such as education, employment status, occupation, time allocation, etc.

In this study, we focus on the nuclear families, i.e., households which consist of two parents and possibly children. This gives us 5541 observations. Some observations do

²The data from the national household survey are not available for use outside SSB.

³The survey does not cover all thirty provinces in China (cities and autonomous districts), mainly due to lack of funding.

not have complete information or have implausible values on the expenditure variables of interest. After deleting these observations, a sample of 5394 households remains which is used for the estimations. Table 1 gives the sample means and standard deviations of the variables we use in the analysis.

We classified the observations into three geographic regions: coastal, middle, and western areas. This distinction has much more than only geographic meaning. The coastal area is historically the most developed and the richest area, while the western area is the least developed and poorest area among the three. According to the sample, the mean log total expenditures of the households in coastal area was 7.90 (about 2697 *yuans*) in 1995, while for the middle and western areas, the numbers were 7.50 (1808 *yuans*) and 7.44 (1703 *yuans*), respectively. Moreover, culture, traditions, and many other aspects of social life are significantly different in the three regions.

Table 1. Variable Definitions and Sample Statistics

Variable	Description	Mean	Std. Dev.
<i>lnc</i>	log total expenditures per capita(<i>yuan</i>)	7.61	0.58
<i>lncinc</i>	log disposable income per capita(<i>yuan</i>)	7.18	0.71
<i>food</i>	food share(%)	54.8	18.7
<i>cereal</i>	cereal share(%)	25.8	13.9
<i>prot</i>	protein share(%)	12.9	8.55
<i>meat</i>	meat share(%)	7.5	7.2
<i>fish</i>	fish share(%)	0.9	1.5
<i>veg</i>	vegetable share(%)	8.4	6.9
<i>alta</i>	alcohol & tobacco share(%)	3.1	2.8
<i>medi</i>	medical share(%)	2.1	4.2
<i>edu</i>	education share(%)	0.7	2.2
<i>num</i>	number of household members	3.97	0.98
<i>padu</i> *)	percentage of adult members (19-)	60.3	21.3
<i>pfadu</i> **)	percentage of female adult members(19-)	29.8	13.1
<i>child6</i>	percentage of children (0-5)	4.3	10.9
<i>girl6</i>	percentage of female children (0-5)	1.9	7.8
<i>child12</i>	percentage of children (6-12)	15.5	19.3
<i>girl12</i>	percentage of female children (6-12)	7.1	12.8
<i>child15</i>	percentage of children (6-12)	7.0	12.0
<i>girl15</i>	percentage of female children (13-15)	3.1	8.2
<i>child18</i>	percentage of children (16-18)	6.9	12.1
<i>girl18</i>	percentage of female children (16-18)	3.2	8.4
<i>pup12</i>	percentage of children at school (6-12)	13.1	18.0
<i>pug12</i>	percentage of female children at school (6-12)	5.9	11.8
<i>pup15</i>	percentage of children at school (13-15)	6.1	11.4
<i>pug15</i>	percentage of female children at school (13-15)	2.6	7.6
<i>pup18</i>	percentage of children at school (16-18)	3.2	8.7
<i>pug18</i>	percentage of female children at school (16-18)	1.4	5.7
<i>pup19</i>	percentage of children at school (19-)	1.0	5.0
<i>pug19</i>	percentage of female children at school (19-)	0.4	3.2
<i>age</i>	average age of husband and wife	41.8	9.38
<i>dcoast</i>	dummy, 1 if household in coastal area	0.317	0.465
<i>dmiddle</i>	dummy, 1 if household in middle area	0.454	0.498
<i>dwest</i>	dummy, 1 if household in western area	0.228	0.420

Note: 5394 observations.

*) : both parents and children older than 18.

**) : spouse and female children older than 18.

Expenditures, self-consumption, and disposable income

The total expenditure variable was constructed from “household total expenditures in cash in 1995” and “self-consumption”, net of “expenditures on house-building and on production goods (such as buying fertilizer or a tractor)”, and net of “taxes”. The reason for subtracting house building expenditures is that house-building has the nature of an investment expenditure rather than consumption. It is an occasional event and is mostly financed by saving (or borrowing). Purchasing of production goods can be viewed as investment behavior and expenditures on these are thus considered as production costs instead of consumption. As shown in Table 1, the mean of log total expenditures per capita in 1995 was 7.61, which, translated into Chinese currency, is 2018 *yuans*.⁴ The average disposable income (which does not include self-consumption) per capita in 1995 was about 1713 *yuans*. To get a clear profile of the distributions and the relation between the two variables, we depict the estimated bivariate kernel density estimates of log total expenditures and log disposable income in Figure 1. The joint distribution is unimodal and shows that the two variables are positively correlated though the correlation is far from perfect (the correlation coefficient is 0.495.).

In 1995, the families’ main outlay was on food, which on average accounted for almost 55% of the family budget. Nearly half of the food expenditures was for cereals, and about 22% for protein intake, which includes the expenditures on meat, fish, eggs, and cooking oil. The expenditure shares on educational goods (excluding tuition fees) is 3.3%. The expenditure share of medical care is 1.8% of total expenditures, which is quite low. The reason is that medical services are heavily subsidized by the government,⁵ implying that for these categories, expenditures as recorded in the survey would substantially underestimate the use of the services.

In rural China, a substantial part of food consumption comes from self-production. Hence, in constructing the food expenditure variable, we have included expenditures in cash as well as the amount reflecting self-consumption. The latter was calculated from the quantity consumed, which was asked in the questionnaire, multiplied with the average price of the good in the province where the household lived.⁶ Table 2 shows the importance of self-consumption for total food consumption in our sample of rural households. On average, self-consumption of food covered 37.6% of total expenditures, and 66.9% of food expenditures. About 90% and 72% of the cereals and vegetables were from self-production. This is consistent with the fact that rural China is still a largely self-contained farmers society. On the other hand, about 72% of the meat and 88% of the fish consumed were from the market.

⁴One US Dollar is about 8.3 yuans.

⁵For example, most of the clinics in rural areas are publicly owned and the prices of their services are stipulated by the government.

⁶In the questionnaire, the households were asked how much money they spent on food. They were also asked, for each commodity, what was the total amount they consumed, how much of this was produced by themselves, and which market price they perceived for this commodity.

Table 2. Distribution of Self-consumption Share(%)

Quantiles	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Mean
Total	13.4	20.4	26.3	32.1	36.9	42.2	47.3	54.1	62.8	37.6
Food	38.0	51.4	60.3	66.6	71.6	75.8	79.7	83.5	88.1	66.9
Cereal	65.8	86.9	94.4	97.7	99.1	99.9	100.0	100.0	100.0	89.6
Meat	0.0	0.0	0.0	0.0	19.0	16.9	43.2	73.5	90.9	27.9
Fish	0.0	0.0	0.0	0.0	0.0	0.0	0.0	10.0	60.0	12.3
Protein	0.0	7.0	17.9	29.2	40.4	52.4	64.2	77.2	89.1	42.7
Vegetable	0.0	40.2	69.5	83.2	89.6	93.9	96.6	98.7	100.0	71.9

Q_j ($j = 1, \dots, 9$ is the j th percentile of the corresponding share.

To show how the size and the structure of the expenditures vary with disposable income, we show the nonparametrically estimated relationships between the log total expenditures and income, between the log expenditures on food and income, and between the proportion of self-consumption (out of the total expenditures) and income in Figures 2, 3, and 4, respectively. Expenditures on consumption are almost constant at low income levels and only increase with income after income reaches a certain threshold level. A more or less similar pattern is obtained for food expenditures. Furthermore, the proportion of expenditures on self-consumption is also almost constant at low income levels before it decreases with income. This implies that the very low income families spend their additional income on activities other than consumption, such as production or investment. Only after income has reached a certain level, families would spend more on consumption and resort more to the markets for food.

Family composition and preliminary evidence of gender bias

Family composition is described by the number of members in certain categories as a percentage of total family size. For example, an adult is defined as someone older than 18. “Percentage of (female) adults” refers then to the number of (female) adults in the family as a percentage of family size. On average, 60% of family members are older than 18 years. All “percentage of (female) children” variables refer to the number of (female) children in the given category, as a percentage of total family size.

From Table 1 we can also see that the average nuclear family in the sample has about four members, i.e. two children. This seems to be in contrast with China’s one child policy. In rural China, however, family planning policy has never been so strict as in urban areas and varies dramatically across regions. Instead, in most places, farmers are normally allowed to have their second child if their first child is a girl or disabled. Also, minorities are also exempted from the policy. In reality, many families even have more than two children, implying that the policy is often not so effective. According to our data (see Table 3 below), only about 26.1% the families have one child, 42.3% of them have two children, and about 20.4% of them have three or more children. There are more boys than girls in the age category below 16, and also in the

category of youngest children. Table 3 shows that there are fewer families without boys than families without girls. Most of the families have one boy, reflecting the fact that in rural China boys are preferred. The main reasons for this preference, among others, are that in a largely farmer's society, men are the main labour providers, and that only the boys are considered as the ones who continue the family roots.

Table 3 also presents some sample statistics for families with one, two, and three children. The one-child families have the highest average income as well as total expenditures per capita. They also have the highest share of protein consumption and lowest share of cereal consumption. This is related to the fact that one-child families are more often in the richer coastal areas than in the poorer innerland, while families with more than two children are more common in the non-coastal areas.

As mentioned in the introduction, it is often found in the literature that female mortality rates are higher than male mortality rates in the same age groups, and that school drop-out rates are higher for girls than for boys. It can also be seen in Table 1 that in all age categories, the school enrollment rates are larger for boys than for girls. These findings can be compared to data from other sources: according to the World Bank, the population female-male ratio for China was 94:100 in 1995. The female gross primary enrollment rate⁷ in 1993 was 116%, against 120% for males. For females, this is the highest enrollment rate in Asia. The secondary school enrollment rate was much lower: 51% for females and 60% for males in 1993.

As a preliminary analysis for this issue, we investigated the probability for a child to be in school using a probit model. We only looked at families with one child in the school attending age category 6 to 18 (there might be additional children in other age groups). Controlling for total family expenditures per capita, age group, geographical area and father's education, we found a positive and significant coefficient for the boy dummy, indicating that boys are more likely to be in school than girls in the same age category.⁸ At the mean of the sample, a boy's probability of being in school is about 7% points higher than a girl's. The estimates of the parameters are presented in Table 4.

⁷Gross enrollment rate is the ratio of total enrollment, regardless of age, to the population of the age group that officially corresponds to the level of education shown. Estimates are based on UNESCO's classification of education levels. Primary, or first level, provides the basic elements of education at elementary or primary school. Source: World Development Indicators 1997, The World Bank.

⁸When we include the interactions between the boy dummy and age, the coefficient of boy dummy becomes insignificant, although still positive. However, the coefficients of the boy dummy and the interaction terms are jointly significant at 5% level. We also tried out several other specifications. For example, we included the second order term of total *Lnc* and/or controlled for children in other age groups. The results remain the same.

Table 3 Sample statistics by Number of Children

No. of children	<i>nchild</i> = 1	<i>nchild</i> = 2	<i>nchild</i> = 3
% of the sample	26.1	42.3	20.4
<u>% given <i>nchild</i></u>			
No boy	36.5	12.1	7.3
One boy	63.5	60.0	47.6
Two boys		27.9	36.3
Three boys			8.8
<u>Sample statistics</u>			
<i>lnexp.</i> (p.c.) (yuans)	7.80(0.58)	7.59(0.56)	7.43(0.52)
<i>lnincome</i> (p.c.) (yuans)	7.34(0.71)	7.15(0.67)	7.02(0.69)
age	40.9(11.7)	40.5(7.9)	42.1(6.7)
coast	0.385(0.49)	0.295(0.46)	0.258(0.44)
middle	0.355(0.48)	0.488(0.50)	0.528(0.50)
west	0.260(0.44)	0.217(0.41)	0.214(0.41)
<u>% of <i>expenditures</i></u>			
food	54.3(19.6)	54.2(18.5)	55.7(17.6)
Cereal	23.0(13.2)	26.0(13.7)	28.9(14.6)
Protein	14.5(9.3)	12.2(8.1)	11.6(7.9)
Meat	8.7(7.9)	7.0(7.0)	6.6(6.3)
Fish	1.0(1.4)	0.8(1.2)	0.8(2.1)
Vegetables	8.4(7.3)	8.2(6.8)	8.4(6.4)
Alco.& tab.	3.6(3.6)	3.1(2.6)	2.8(2.5)
Medical	2.3(4.8)	2.0(3.6)	2.0(4.4)
Education	0.5(1.9)	0.8(2.4)	0.7(1.8)

In parentheses are standard deviations.

Table 4. Estimates of the probit model

Variable	Estimates	<i>t</i> -values
<i>Lnc</i>	0.124	(2.02)
<i>Dummy boy</i>	0.199	(2.99)
<i>Dcoast</i>	-0.162	(-1.69)
<i>Dmiddle</i>	-0.319	(-3.56)
<i>Dummy age between 6 and 12</i>	1.077	(13.88)
<i>Dummy age between 13 and 15</i>	1.287	(13.00)
<i>Father's education (> 6 years)</i>	0.458	(3.38)
<i>Father's education (≥ 4 and ≤ 6 years)</i>	0.310	(2.23)
<i>Constant</i>	-1.611	(-3.37)

Note: 1644 observations.

4 Results

The Shape of the Engel Curves

Figure 5 gives the parametrically and semiparametrically estimated Engel curves together with 95% uniform confidence bands (see Härdle, 1990 for reference), where the endogeneity of total expenditures is controlled for. For some goods, such as cereals, alcohol and tobacco, medical care, and education, the linear model gives a reasonable description of the Engel curve relation. But for other goods, such as food, meat, fish, and vegetables, the linear model fits the data poorly and a quadratic model would fit better.⁹ For these goods, a linear model would lead to biased predictions for low and/or high income households. For example, a linear model would predict a too large negative income slope of the food share for the low income households.

The findings here are different from those in Blundell *et al.* (1998). Their findings based on UK data suggest that food share Engel curves are linear. A reason for the difference could be the different real income level of the low income households in our sample. Low income families are the poorest among farmers who are already the poor in China. In 1995, an average farmer consumes only about 2018 *yuan*s per year, which is less than US\$250. For these very poor families, the percentage of expenditures on food remained relatively constant as income increased because even with income increased, some of them were still barely able to feed themselves. Only after income has reached a certain level, the food share decreases, and additional resources are more and more allocated to non food consumption. Bhalotra and Attfield (1998) find a similar result for rural Pakistan, where food consumption is 52% of the household budget, on average, compared to 54.8% in our sample. In developed countries, there are hardly any “absolutely poor” households, and the average food share is much smaller (34% to 37% in the UK data used by Blundell *et al.*, 1998).

For the meat and the fish shares, the linear model Engel curve assumption is rejected even more clearly. For low income households, meat and fish are luxuries, and the budget share of meat consumption increases with income. At higher income levels however, less of the additional resources is allocated to fish, and the share spent on meat decreases with income.

In Table 5 we present the elasticities of consumption expenditures with respect to total expenditures, as computed by both the semiparametric and the linear model, where endogeneity of total expenditures is corrected for. It shows that the elasticities according to the linear and semiparametric model are similar for most goods. Only for cereals we find a substantial difference. The similarity rests on the fact that the elasticities we calculated are ‘elasticities of the average’ –not surprisingly, the two models give similar predictions for the means. According to the elasticities, fish is a luxury good, while food, cereals, vegetables, protein, and education are necessities. The other three elasticities

⁹For these goods, the estimates of the quadratic model are all within the confidence bands, so that a quadratic model would not be rejected against the semiparametric model.

are close to one, and not significantly different from one in either model.

Thus we find that households with higher income allocate less and less of their additional income to food consumption. Moreover, they spend more on food containing more energy. If total expenditures increase, cereal consumption decreases relative to total food consumption. For a one percent increase of total expenditures, the household increases food consumption by 0.61% and cereal consumption only by 0.22%. At the same time, the richer households spend more of their food budget on meat, fish and other types of food which are rich of protein.

Table 5. Estimated Income Elasticities

<i>Goods</i>	Semiparametric Model	Linear Model
<i>Food</i>	0.559 [0.524, 0.600]	0.572 [0.540, 0.606]
<i>Cereal</i>	0.232 [0.172, 0.295]	0.168 [0.104, 0.240]
<i>Protein</i>	0.881 [0.821, 0.964]	0.942 [0.869, 1.018]
<i>Meat</i>	0.983 [0.903, 1.092]	1.068 [0.973, 1.168]
<i>Fish</i>	1.819 [1.689, 1.963]	1.949 [1.814, 2.088]
<i>Vegetables</i>	0.496 [0.401, 0.590]	0.524 [0.415, 0.625]
<i>Alcohol & Tobacco</i>	0.991 [0.900, 1.076]	0.978 [0.871, 1.090]
<i>Medical care</i>	0.976 [0.702, 1.234]	1.009 [0.760, 1.248]
<i>Education</i>	0.606 [0.224, 0.949]	0.519 [0.415, 0.625]

In brackets are 95% confidence intervals.

In both models endogeneity of total expenditures is corrected for.

In table 6, we illustrate the influence of the correction for endogeneity of total expenditures by comparing the estimated elasticities from the model with and without the correction. $\hat{\rho}_i$ is the estimates of ρ_i in equation (11). A significant value of $\hat{\rho}_i$ implies that exogeneity is rejected. For five out of nine goods the Engel curve exhibits significant endogeneity of expenditures. For example, a significant positive sign for $\hat{\rho}_i$ is found for cereals, the clearest necessity among the goods. In the other cases where endogeneity is significant, the correlation between the errors is negative.¹⁰ Comparing the results of Model I in Table 6 with the confidence intervals in Table 5, we find that for the goods (cereals, protein, meat, fish, and alcohol and tobacco) where the endogeneity plays a significant role, the elasticities are biased if endogeneity is not taken into account. The direction of the bias depends upon the sign of the correlation –the elasticities will be underestimated if the correlation is negative.

¹⁰This could be explained by the fact that a positive measurement error on total expenditures induces a negative measurement error on the share.

Table 6. Estimated Income Elasticities and Endogeneity

<i>Commod.</i>	Semiparametric Models			Linear Models		
	Model I	Model II		Model I	Model II	
	<i>elas.</i>	<i>elas.</i>	$\hat{\rho}_i$	<i>elas.</i>	<i>elas.</i>	$\hat{\rho}_i$
<i>Food</i>	0.574	0.559	0.009(0.93)	0.598	0.572	0.014(1.48)
<i>Cereal</i>	0.408	0.232	0.046(6.18)	0.355	0.168	0.044(6.21)
<i>Protein</i>	0.700	0.881	-0.028(-5.34)	0.769	0.942	-0.024(-4.84)
<i>Meat</i>	0.818	0.983	-0.015(-3.47)	0.914	1.068	-0.128(-3.15)
<i>Fish</i>	0.947	1.819	-0.011(-11.6)	1.085	1.949	-0.010(-11.3)
<i>Veget.</i>	0.556	0.496	0.007(1.59)	0.628	0.524	0.009(2.08)
<i>Alc.&Tab.</i>	0.763	0.991	-0.010(-5.48)	0.702	0.978	-0.009(-5.40)
<i>Medical</i>	0.989	0.976	0.000(0.12)	1.037	1.009	0.001(0.31)
<i>Education</i>	0.889	0.606	0.003(1.81)	0.830	0.519	0.002(1.66)

Model I: no correction for endogeneity of total expenditures.

Model II: endogeneity of total expenditures is corrected for.

t-values in parentheses.

Sexual Bias and Consumption Patterns

The estimates of the parameter vectors β_i in the partially linear model are shown in Table 7. We focus on consumption differences related to differences in the number of male and female household members in the same age categories. These differences are reflected by the coefficients of the variables relating to the numbers of female members (as a percentage of family size).

A negative sign of such a coefficient for some given commodity means that a family with a female member in this age group consumes less of this commodity than a comparable family with a male member in the same age group. Most of these variables are insignificant for the various food categories, but there are some exceptions. The significantly negative coefficients of *girl15* for food and cereal suggest that girls aged at 12 to 15 consume less of this. On the other hand, for the youngest age group, the protein intake of girls is significantly larger than for boys.

For a typical adult good category alcohol and tobacco, we find insignificant differences between boys and girls in the younger age groups. Since alcohol and tobacco are the only typical “adult goods”, this implies that there is no evidence of sexual bias according to the Deaton (1989) method.

Thus in the food and adult goods equations, no convincing evidence of sexual bias can be found. This conclusion largely coincides with most of the findings for various countries in the literature (for example, Deaton (1989); see also Deaton (1997) for a review). It does not say that there is no sexual bias, but only that it is not revealed by food consumption patterns.

For educational goods, we have defined the family composition variables in a different way. Instead of the total number of boys and girls in a given age group, we consider the

number of boys and girls attending school. We also include the age group older than 18. Students in this age group mainly attend post-secondary school (college education, for example). Thus the estimates are conditional on the enrollment decisions. Differences in enrollment between boys and girls are discussed in Section 3. We find that educational expenditures for boys and girls do not differ significantly for the younger age groups. For the older age groups (16-18 and older than 18), however, less is spent for girls than for boys. Thus other than the food equations, there is some evidence of bias against girls in the educational goods equation. This is further enforced by the fact that school enrollment rates are higher for boys than for girls (see Table 4).

Some differences between expenditure patterns in families with male versus female adults can be seen from the coefficients on “percentage of female adults.” Since we also include the “percentage of adults,” this can be interpreted in the same way as the percentage of girls variables. The coefficients of *fadu* for cereals and fish are negative, while that for medical expenditures is positive. This suggests that female workers consume less cereal and fish than male adults.

Table 7 : Estimates of the partially linear model

β_i	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>age</i> /10 ²	-0.592 (-2.64)	-0.026 (-0.16)	-0.277 (-2.36)	-0.306 (-3.18)	-0.015 (-0.71)	-0.246 (-2.46)	0.052 (1.25)	0.111 (1.75)	0.027 (0.96)
<i>age</i> 2/10 ⁴	0.695 (2.68)	0.150 (0.78)	0.274 (2.02)	0.319 (2.87)	0.024 (0.94)	0.271 (2.33)	-0.099 (-2.07)	-0.083 (-1.12)	-0.037 (-1.18)
<i>dcoast</i>	-0.009 (-1.37)	-0.027 (-5.40)	-0.050 (-14.0)	-0.066 (-22.6)	0.008 (12.4)	0.037 (12.1)	0.009 (7.04)	0.002 (0.99)	0.002 (2.30)
<i>dmiddle</i>	-0.065 (-12.3)	-0.025 (-6.29)	-0.069 (-25.1)	-0.080 (-35.5)	0.005 (9.23)	0.018 (7.70)	0.005 (4.78)	0.003 (1.75)	0.000 (0.60)
<i>num</i>	-0.083 (-6.82)	-0.034 (-3.71)	-0.027 (-4.30)	-0.008 (-1.61)	0.003 (2.39)	-0.012 (-2.24)	-0.008 (-3.41)	-0.012 (-3.41)	-0.000 (-0.04)
<i>padu</i>	0.036 (-0.16)	0.035 (2.10)	-0.015 (-1.31)	-0.005 (-0.47)	-0.003 (-1.24)	-0.017 (-1.67)	0.001 (0.36)	-0.023 (-3.56)	-
<i>pfadu</i>	-0.020 (-0.87)	-0.035 (-2.09)	0.007 (0.60)	0.005 (0.53)	-0.005 (-2.47)	0.005 (0.45)	-0.003 (-0.66)	0.023 (3.61)	-
<i>child6</i>	-0.047 (-1.08)	0.061 (1.88)	-0.071 (-3.15)	-0.046 (-2.48)	0.001 (0.16)	-0.029 (-1.50)	-0.020 (-2.52)	0.018 (1.43)	0.002 (0.32)
<i>girl6</i>	0.001 (0.03)	-0.018 (-0.61)	0.043 (2.08)	0.025 (1.51)	0.001 (0.14)	-0.010 (-0.55)	0.004 (0.52)	-0.000 (-0.02)	-0.005 (-0.88)
<i>child12*</i>	-0.036 (-1.08)	0.045 (1.81)	-0.042 (-2.39)	-0.025 (-1.78)	-0.001 (-0.16)	-0.013 (0.91)	-0.017 (-2.79)	0.005 (0.48)	0.006 (2.14)
<i>girl12*</i>	-0.005 (-0.20)	-0.004 (-0.27)	-0.003 (-0.24)	-0.006 (-0.60)	-0.000 (-0.22)	-0.005 (-0.46)	-0.000 (-0.08)	0.004 (0.69)	-0.001 (-0.31)
<i>child15*</i>	-0.028 (-0.81)	0.079 (3.08)	-0.064 (-3.56)	-0.041 (-2.79)	-0.004 (-1.19)	-0.017 (-1.11)	-0.013 (-1.99)	-0.010 (-1.07)	0.014 (3.89)
<i>girl15*</i>	-0.089 (-2.79)	-0.067 (-2.81)	-0.008 (0.49)	0.012 (0.88)	-0.003 (-0.93)	0.013 (0.90)	-0.011 (-1.83)	0.004 (0.49)	-0.001 (-0.11)
<i>child18*</i>	-0.059 (-1.73)	0.056 (2.19)	-0.053 (-2.97)	-0.029 (-1.96)	-0.007 (-2.00)	-0.018 (-1.16)	-0.020 (-3.22)	-0.010 (-1.07)	0.024 (5.14)
<i>girl18*</i>	0.002 (0.08)	-0.026 (-1.10)	0.028 (1.68)	0.017 (1.25)	-0.002 (-0.56)	0.003 (0.45)	-0.000 (-0.07)	0.014 (1.59)	-0.018 (-2.58)
<i>pup19</i>	-	-	-	-	-	-	-	-	0.048
<i>pug19</i>	-	-	-	-	-	-	-	-	(6.05)
	-	-	-	-	-	-	-	-	-0.039
	-	-	-	-	-	-	-	-	(-3.18)
ρ_i	0.009 (0.93)	0.046 (6.18)	-0.028 (-5.34)	-0.015 (-3.47)	-0.011 (-11.6)	0.007 (1.59)	-0.010 (-5.48)	0.000 (0.12)	0.003 (1.81)

Note: 1. (1): Food; (2): Cereal; (3): Protein; (4): Meat; (5): Fish; (6): Vegetable;

(7) Alcohol and tobacco; (8): Medical; (9): Education.

2. t -values are in parentheses

3. All the estimates allow for endogeneity of total expenditures. See Table A in the appendix for estimates under the assumption that total expenditure is exogenous.

* For education, the variables with “*” are replaced by the percentages of children attending school in the same age groups, *e.g.*, *child12* and *girl12* are replaced by *pup12* and *pug12*, etc.

To test the sexual bias formally, we conducted two Wald tests for each commodity, the results of which are in Table 8. One is to test the null hypothesis that all the coefficients of female-member-related variables are zero –there is no bias between boys and girls for all the age groups. For cereals, medical care, and education, the null hypothesis is rejected. The other is to test the null hypothesis that the coefficients of all female-member-related variables are the same—the bias is the same for all the age groups. This null hypothesis is rejected only for education.

The results in Table 7 give some interesting information beyond sexual bias. First, the fact that all but one of the coefficients for household size (the variable *num*) are negative, suggests economies of scale in household consumption. A similar result was found by Bhalotra and Attfield (1998) for rural Pakistan. Second, households in coastal and middle regions, *ceteris paribus*, spend less than those in the western regions on cereals and meat, but more on fish, vegetables, education, alcohol and tobacco. The differences might well be due to price differentials between regions. For example, in coastal area, the average price for fish reported by the households is about 7.5 *yuans*, but it is about 8.4 *yuans* in western area. Finally, the endogeneity does not seem to alter the estimates of the family composition parameters systematically. In particular, the signs and the significance of the estimates are largely unchanged.

The estimates for the model without correction for the endogeneity, are presented in Table A in the appendix. They are not very different from those in Table 7. Thus ignoring endogeneity biases the elasticity estimates but not the conclusions on the other variables of interest or the conclusions on sexual bias.

Table 8. Specification tests

$H_0 :$	no sexual bias	same degree of bias across age
Food	8.52	6.21
Cereal	13.22*	5.34
Protein	7.70	5.31
Meat	5.33	3.99
Fish	7.09	3.43
Vegetable	1.59	1.54
Alco.& tab.	4.05	3.23
Medical	15.54*	6.49
Education	16.88*	12.89*
	$\chi^2_{.050}(5) = 11.07$	$\chi^2_{.050}(4) = 9.49$

* H_0 is rejected at 5% level

5 Conclusions

We analyzed expenditure data from a cross-section household survey in rural China. We compared parametric and semiparametric estimates of demand curves for nine expenditure categories: food and several food categories, education, and medical expenses.

Expenditure shares on these categories are explained from total expenditures, family composition, and regional dummies. Expenditures include self-consumption of home produced goods, which represents a substantial share of total consumption in our sample.

It is likely that total expenditures is jointly determined with expenditure shares. We therefore account for endogeneity of total expenditures, applying the ‘augmented regression approach’. It appears that for many goods, total expenditures is significantly endogenous, and failure in correcting for endogeneity leads to substantial biases in some of the elasticity estimates.

We compare the linear Engel curve specification with a semiparametric partially linear model, in which total expenditures enters in a flexible way, while the influence of demographics etc. remains linear. Our semiparametric estimates of Engel curves suggest that Linear Engel curves (or PIGLOG form Engel curves) are not suitable for most of the commodity categories we considered. Particularly in case of food, the linear model is clearly rejected by our data. This is in contrast to findings for the developed countries. The reason could be that our data contain a large group of very low income households.

We find significant indications for economies of scale in consumption of seven out of nine goods. The variation in consumption behavior among different geographic regions is also clearly shown by the results.

We include dummies for the numbers of female children and adults in various age groups to test for sexual bias. Hardly any discriminating effects between boys and girls are found in the food consumption patterns. Probits for school enrollment suggest that boys have a larger probability to go to school than girls of the same age. Moreover, conditional on enrollment, there is evidence that educational expenditures for boys in the age groups 16 and older are larger than for girls in the same age group. We find that the presence of adult female workers in the family significantly increases expenditures for medical care, which might imply that the health of these females is neglected. These findings confirm the documented findings for other developing countries that the sexual bias often comes from other sources such as higher mortality rates and lower school enrollment rates among females but not through consumption of common goods.

References

- Ahmad, A. and J. Morduch (1993), "Identifying Sex-bias in the Allocation of Household Resources: Evidence from Linked Household Surveys from Bangladesh," Harvard Institute of Economic Research Discussion Paper 1636, Cambridge: Harvard University.
- Arnold, F. and Z. Liu (1986), "Sex Preference, Fertility and Family Planning in China," *Population and Development Review*, 12(2), 221-246.
- Atkinson, A., J. Gomulka, and N. Stern (1990), "Spending on Alcohol: evidence from the Family Expenditure Survey 1970-1983," *Economic Journal*, 100, 808-827.
- Banks, J., R. Blundell, and A. Lewbel (1997), "Quadratic Engel Curves and Consumer Demand," *The Review of Economics and Statistics*, 79, 527-539.
- Beran, R. (1988), "Prepivoting Test Statistics: A Bootstrap View of Asymptotic Refinements," *Journal of the American Statistical Association*, 83, 687-697.
- Blundell, R., A. Duncan, and K. Pendakur (1998), "Semiparametric Estimation and Consumer Demand," *Journal of Applied Econometrics*, 13(5), 435-461.
- Blundell, R., P. Pashardes, and G. Weber (1993), "What Do We Learn about Consumer Demand Patterns from Micro-data?," *American Economic Review*, 83, 570-597.
- Bhalotra, S. and C. Attfield (1998), "Intrahousehold Resource Allocation in Rural Pakistan: A Semiparametric Analysis," *Journal of Applied Econometrics*, 13(5), 435-461.
- Browning, M. and R. Subramaniam (1995), "Gender-bias in India: Parental Preferences or marriage costs," McMaster University, mimeograph.
- Burgess, R. and J. Zhuang (1998), "Gender Biases in Intrahousehold Allocation in Rural China," STRICERD Development Economics Research Programme, London School of Economics, mimeograph.
- Craven, P. and G. Wahba (1979), "Smoothing Noisy Data with Spline Function: Estimating the Correct Degree of Smoothing by the Method of Generalized Cross-Validation," *Numerische Mathematik*, 31, 377-403.
- Das Gupta, M. (1987), "Selective Discrimination Against Female Children in Rural Punjab, India," *Population and Development Review*, 13(1), 77-100.
- Deaton, A. (1989), "Looking for Boy-girl discrimination in Household Expenditure data," *World Bank Economic Review*, 3, 1-15.
- Deaton, A. (1997), *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*, Baltimore: Johns Hopkins University

- Press.
- Deaton, A., and J. Muellbauer (1980), "An Almost Ideal Demand System," *American Economic Review*, 70, 312-326.
- D'Souza, S. and L. Chen (1980), "Sex differentials in mortality in rural Bangladesh," *Population and Development Review*, 6(2), 257-270.
- Fan, S., G. Cramer, and E. Wailes (1994), "Food demand in Rural China: evidence from rural household survey," *Agricultural Economics* 11, 61-69.
- Fan, S., E. Wailes, and G. Cramer (1995), "Household demand in Rural China: A Two-Stage LES-AIDS Model," *American Journal of Agricultural Economics* 77, 54-62.
- Gorman, W. (1981), "Some Engel Curves," in *Essays in the Theory and Measurement of Consumer Behaviour*, A. Deaton (ed.), Cambridge, UK: Cambridge University Press.
- Härdle, W., (1990), "Applied Nonparametric Regression," Econometric Society Monographs 19, Cambridge University Press.
- Härdle, W. and M. Jerison (1991), "Cross-Sectional Engel Curves over time," *Recherches Economiques de Louvain*, 57, 391-431.
- Hausman, J., W. Newey, H. Ichimura, and J. Powell (1991), "Identification and Estimation of Polynomial Errors in Variables Models," *Journal of Econometrics*, 50, 273-296.
- Hausman, J., W. Newey, and J. Powell (1995), "Nonlinear Errors in Variables: estimation of some Engel curves," *Journal of Econometrics*, 65, 205-234.
- Holly, A. and J. Sargan (1982), "Testing for Exogeneity in a Limited Information Framework," *American Economic Review*, 70, 268-272.
- Horowitz, J. (1997), "Bootstrap Methods in Econometrics: Theory and Numerical Performance," in *Advances in Economics and Econometrics: Theory and Applications*, Vol. 3, ed. by D. Kreps and K. W. Wallis. Cambridge: Cambridge University Press, 188-222.
- Horowitz, J. (1998), "Bootstrap Methods For Median Regression Models," *Econometrica*, 66(6), 1327-1351.
- Jorgenson, D., L. Lau, and T. Stoker (1980), "Welfare Comparison and Exact Aggregation," *American Economic Review*, 70, 268-272.
- Leser, C. (1963), "Forms of Engel Curves," *Econometrica*, 31, 694-703.
- Lee, M.J. (1996), *Methods of Moments and Semiparametric Econometrics for Limited Dependent Variables Models*, Springer-Verlag New York.
- Lewbel, A. (1991), "The Rank of Demand Systems: theory and nonparametric estimation," *Econometrica*, 94, 979-1000.
- Mack, Y. and H. Müller (1989), "Derivatives estimation in nonparametric regression with random predictor variable," *Sankhya*, A51, 59-72.
- Newey, W., J. Powell and J. Walker (1990), "Semiparametric estimation of

- selection models: some empirical results," *American Economic Review Proceedings*, 80, 324-328.
- Robinson, P. (1988), "Root-N-Consistent Semiparametric Regression," *Econometrica*, 56, 931-954.
- Stoker, T. (1991), *Lectures on Semiparametric Econometrics*, CORE Lecture Series, CORE Foundation, Universite Catholique de Louvain.
- Subramaniam, R. (1996), "Gender-bias in India: The Importance of Household Fixed Effects," *Oxford Economic Papers*, 48(2), 280-299.
- Subramaniam, R. and A. Deaton (1991), "Gender Effects in Indian Consumption Patterns," Princeton University, mimeograph.
- Working, H. (1943), "Statistical Laws of Family Expenditure," *Journal of the American Statistical Association*, 38, 43-56.
- Zeng, Y. (1988), "Changing Demographic Characteristics and the Family Status of Chinese Women," *Population Studies*, 42, 183-203.
- Zeng, Y., P. Tu, B. Gu, Y. Xu, B. Li, and Y. Li (1993), "Causes and Implications of the Recent Increase in the Reported Sex Ratio at Birth in China," *Population and Development Review*, 19(2), 283-302.

Appendix

**Table A: Estimates of the partially linear model
(no correction for endogeneity)**

β_i	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>age/10²</i>	-0.593 (-2.64)	-0.031 (-0.19)	-0.274 (-2.33)	-0.305 (-3.16)	-0.014 (-0.64)	-0.247 (-2.47)	0.053 (1.27)	0.111 (1.75)	0.023 (0.80)
<i>age2/10⁴</i>	0.701 (2.70)	0.178 (0.92)	0.257 (1.89)	0.310 (2.78)	0.017 (0.66)	0.275 (2.37)	-0.105 (-2.19)	-0.082 (-1.12)	-0.030 (-0.98)
<i>dcoast</i>	-0.013 (-2.14)	-0.043 (-9.89)	-0.040 (-13.0)	-0.061 (-24.1)	0.012 (21.0)	0.034 (13.1)	0.012 (11.4)	0.002 (1.07)	0.001 (1.62)
<i>dmiddle</i>	-0.065 (-12.5)	-0.028 (-7.18)	-0.067 (-24.5)	-0.079 (-35.3)	0.006 (10.8)	0.018 (7.55)	0.005 (5.57)	0.003 (1.75)	0.000 (0.37)
<i>num</i>	-0.079 (-6.99)	-0.012 (-1.46)	-0.040 (-6.83)	-0.015 (-3.16)	0.002 (2.18)	-0.009 (-1.77)	-0.012 (-5.92)	-0.012 (-3.64)	0.001 (0.83)
<i>padu</i>	0.057 (-0.23)	0.025 (1.51)	-0.009 (-0.80)	-0.001 (-0.14)	-0.000 (-0.13)	-0.018 (-1.83)	0.004 (0.88)	-0.023 (-3.59)	-
<i>pfadu</i>	-0.019 (-0.84)	-0.032 (-1.87)	0.005 (0.38)	0.004 (0.42)	-0.006 (-2.82)	0.005 (0.50)	-0.004 (-0.84)	0.024 (3.62)	-
<i>child6</i>	-0.044 (-1.01)	0.077 (2.38)	-0.081 (-3.58)	-0.051 (-2.46)	-0.003 (-0.77)	-0.027 (-1.38)	-0.024 (-2.96)	0.018 (1.44)	0.003 (0.55)
<i>girl6</i>	0.001 (0.03)	-0.017 (-0.59)	0.042 (2.06)	0.025 (1.50)	0.000 (0.10)	-0.010 (-0.54)	0.004 (0.51)	-0.000 (-0.02)	-0.005 (-0.87)
<i>child12*</i>	-0.035 (-1.05)	0.050 (2.03)	-0.045 (-2.58)	-0.027 (-1.91)	-0.002 (-0.57)	-0.013 (0.85)	-0.018 (-2.98)	0.005 (0.49)	0.007 (2.38)
<i>girl12*</i>	-0.004 (-0.19)	-0.003 (-0.19)	-0.003 (-0.30)	-0.006 (-0.64)	-0.001 (-0.35)	-0.004 (-0.45)	-0.001 (-0.15)	0.004 (0.69)	-0.001 (-0.30)
<i>child15*</i>	-0.028 (-0.81)	0.080 (3.10)	-0.065 (-3.57)	-0.041 (-2.80)	-0.004 (-1.21)	-0.017 (-1.11)	-0.013 (-2.00)	-0.010 (-1.07)	0.014 (3.99)
<i>girl15*</i>	-0.091 (-2.83)	-0.072 (-3.03)	-0.011 (0.69)	0.014 (1.00)	-0.002 (-0.51)	0.012 (0.84)	-0.010 (-1.63)	0.004 (0.48)	-0.001 (-0.17)
<i>child18*</i>	-0.061 (-1.79)	0.046 (1.80)	-0.047 (-2.63)	-0.026 (-1.75)	-0.004 (-1.26)	-0.019 (-1.26)	-0.018 (-2.88)	-0.010 (-1.08)	0.024 (5.08)
<i>girl18*</i>	0.003 (0.10)	-0.023 (-0.95)	0.026 (1.55)	0.016 (1.17)	-0.003 (-0.82)	0.003 (0.24)	-0.001 (-0.20)	0.015 (1.59)	-0.018 (-2.54)
<i>pup19</i>	-	-	-	-	-	-	-	-	0.048 (6.02)
<i>pug19</i>	-	-	-	-	-	-	-	-	-0.039 (-3.20)

Note: 1. (1): Food; (2): Cereal; (3): Protein; (4): Meat; (5): Fish; (6): Vegetable;
(7) Alcohol and tobacco; (8): Medical; (9): Education.

2. *t*-values are in parentheses

* For education, the variables with ‘*’ are replaced by the percentages of children attending school in the same age groups, *e.g.*, *child12* and *girl12* are replaced by *pup12* and *pug12*, etc.

Figure 1 : Estimated Bivariate Kernel density of total expenditures and disposable income per capita

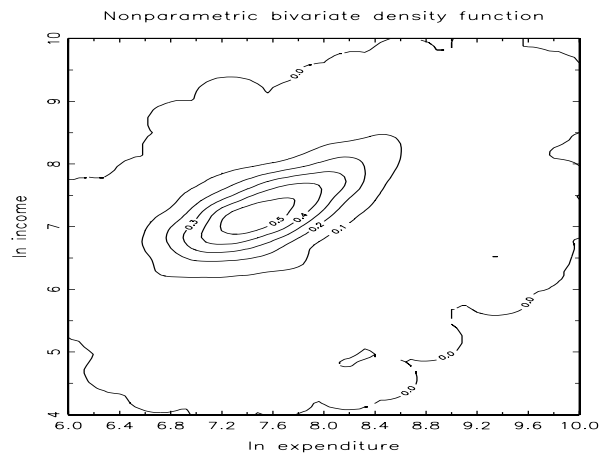


Figure 2 : Nonparametrically estimated relationship between total expenditures and disposable income per capita

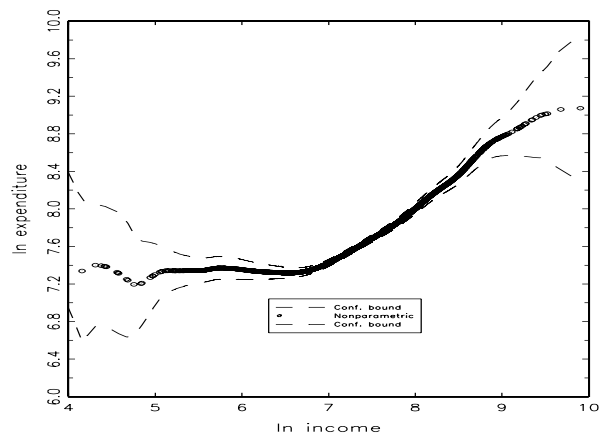


Figure 3 : Nonparametrically estimated relationship between food expenditures and disposable income per capita

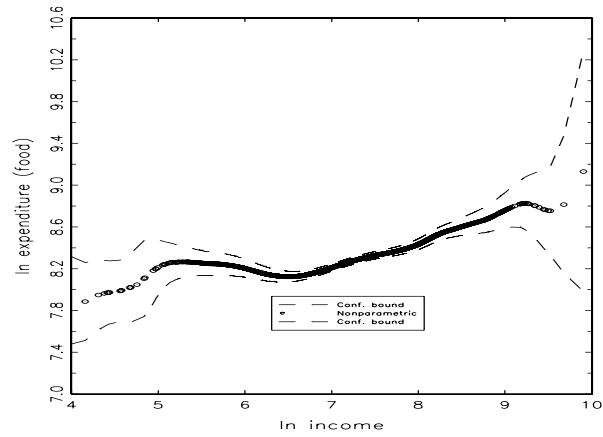


Figure 4 : Semiparametrically estimated relationship between proportion of self-consumption and disposable income per capita

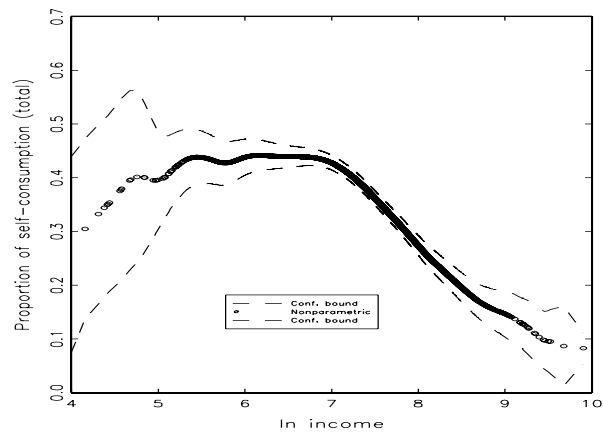


Figure 5 : Nonparametric vs Parametric Engel Curves

